

学校编码: 10384

分类号_____密级_____

学号: 23020071151315

UDC_____

厦门大学

硕士学位论文

哼唱识别及音乐检索系统的研究

Research on Humming Recognition and
Music Retrieval System

钟声声

指导教师姓名: 冯寅 副教授

专业名称: 计算机应用技术

论文提交时间: 2010 年 5 月

论文答辩日期: 2010 年 月

学位授予日期: 2010 年 月

答辩委员会主席:

陈国龙

评阅人:

2010 年 5 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名): 钟奇志

2010年5月29日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（ ） 1. 经厦门大学保密委员会审查核定的保密学位论文，
于 年 月 日解密，解密后适用上述授权。

（ ） 2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）： / 钟齐志

2010年5月29日

摘 要

现代科技高速发展，人们处于一个信息爆炸的时代。通过互联网，人们可以接触到无穷无尽的信息。这些信息也包括一门古老的艺术——音乐。为了找寻需要的歌曲，人们需要一项专注于音乐信息的检索技术。检索需要用户提供检索关键字。显然，哼唱是最容易推广的音乐检索关键字的输入方法。

哼唱旋律识别技术是将哼唱文件通过一定的算法识别成音符的音高和时值序列的技术。音乐检索系统是用户通过一定的方式，将音乐的节奏、旋律、音色等信息输入计算机，计算机通过这些信息从音乐数据库中检索出与这些信息匹配的目标歌曲的系统。将这两者结合可以称为基于哼唱旋律识别技术的音乐检索系统。本文为实现哼唱旋律识别检索系统，分别针对哼唱旋律识别技术和对应的音乐检索系统做出深入研究。

在哼唱旋律识别方面，本文首先提出了基于能量的多层次音块切割方法。该切割方法对哼唱音符切换时的能量变化的不同情况，采用不同方法处理。然后，本文提出了倍音列音高计算模型。倍音列理论的使用，可以在一定程度上提高音高识别的准确性。最后，论文提出了基于调式理论的音高规整方法。该方法用于改进音高识别算法的准确性。在音乐检索系统方面，本文设计了基于哼唱旋律识别结果的数据库建立方案和音乐检索算法。

本文根据以上两方面的研究成果，设计和开发了一个基于哼唱旋律识别技术的音乐检索系统。该系统除了使用了以上提到的理论，还引入了一种人机交互手段，用以提高哼唱旋律识别的正确性和哼唱检索结果的命中率。

关键词：哼唱旋律识别；音乐检索；人机交互

厦门大学博硕士论文摘要库

Abstract

With the rapid development of modern science and technology, people are in an era of information explosion. Through the Internet, people have access to enormous information, which also includes an ancient art - music. In order to find songs people want, they need a retrieval technology focused on music information. Such a retrieval requires users to provide keywords. Obviously, humming is the best music retrieval keyword input method which is most easily popularized.

Humming melody recognition technology is that humming documents are identified into a sequence of notes of the pitch and time value by certain algorithms. Music retrieval system is that the user put the music's rhythm, melodies, sounds and other data into the computer through a certain way, and then the computer retrieves the target songs that match the information from the music database. The combination can be called music retrieval system based on humming melody recognition technology. This dissertation aims for achieving the humming melody recognition and retrieval system, it studies the humming melody recognition and the corresponding music retrieval system respectively.

In the aspect of humming melody recognition, this dissertation proposed a multi-level tone cutting method based on the energy block. The cutting method adopts different approach according to different situations of energy changes of singing notes switch. Then, this paper puts forward overtone series pitch calculation model. the use of overtone series theory can improve the accuracy of pitch recognition to a certain extent. Finally, the dissertation proposes an adjusting approach based on the tuning pitch theory, which is used to improve the accuracy of pitch recognition algorithm. In the aspect of music retrieval system, the database establishing scheme and the music retrieval algorithm are designed based on the humming melody recognition result.

According to two aspects of this research above, this dissertation designs and develops a music retrieval system based on humming melody recognition technology.

The system also introduced a means for human-computer interaction to improve the accuracy of humming melody recognition and the hit rate of humming retrieval results in addition to using the above-mentioned theory.

Key Words: Humming melody recognition; Music retrieval; Human-Computer Interaction

厦门大学博硕士论文摘要库

目 录

中文摘要	I
英文摘要	III
第一章 绪论.....	1
1.1 哼唱识别与音乐检索系统概述	1
1.2 研究的目的和意义	1
1.2.1 研究的目的	1
1.2.2 研究的意义	1
1.3 相关研究概况和问题的提出	2
1.4 本论文创新点	4
1.5 论文篇章结构	5
第二章 哼唱旋律识别技术.....	7
2.1 音块切分技术	7
2.1.1 概述	7
2.1.2 基于能量的音块划分	8
2.1.3 其它划分方法探讨	11
2.1.4 小结	13
2.2 音高识别技术	14
2.2.1 概述	14
2.2.2 离散时间傅立叶变换理论	14
2.2.3 倍音列音高计算模型	23
2.2.4 小结	32
2.3 音高规整技术	33
2.3.1 概述	33
2.3.2 调、调号和音乐调式理论	33
2.3.3 基于调式的音高规整技术	36
2.4 本章小结	42

第三章 哼唱检索系统设计探讨	45
3.1 概述.....	45
3.2 音乐数据库的建立.....	46
3.2.1 乐谱特征格式	46
3.2.2 数据库文件格式.....	50
3.3 检索算法	52
3.4 本章小结	54
第四章 系统的实现和实验结果	55
4.1 系统功能设计	55
4.2 人机交互功能	58
4.2.1 音块划分	58
4.2.2 调音过程	61
4.2.3 音高规整	62
4.3 实验结果及分析.....	62
4.3.1 哼唱旋律识别实验.....	62
4.3.2 哼唱检索实验	65
第五章 总结与展望.....	67
5.1 全文总结	67
5.2 今后的工作	68
附 录	69
参 考 文 献.....	71
攻读硕士期间发表的论文.....	75
致 谢	77

Contents

Abstract in Chinese	I
Abstract in English	III
Chapter 1 Introduction	1
1.1 Overview of humming melody recognition and retrieval system	1
1.2 Purpose and significance of the dissertation	1
1.2.1 Purpose of the dissertation	1
1.2.2 Significance of the dissertation.....	1
1.3 Related research and Presentation of questions	2
1.4 Innovation	4
1.5 Structure of the dissertation	5
Chapter 2 Humming melody recognition technology	7
2.1 Tone block cutting technology	7
2.1.1 Overview	7
2.1.2 Tone block cutting based on the energy.....	8
2.1.3 Other methods.....	11
2.1.4 Summary	13
2.2 Pitch recognition technology	14
2.2.1 Overview	14
2.2.2 Discrete fourier transform theory	14
2.2.3 Pitch calculation model based on overtone series theory.....	23
2.2.4 Summary	32
2.3 Pitch trim technology	33
2.3.1 Overview	33
2.3.2 Tune, key signature and mode theory	33
2.3.3 Pitch trim technology based on tune theory	36
2.4 Summary	42

Chapter 3 Humming retrieval system design	45
3.1 Overview	45
3.2 Music database	46
3.2.1 Music character format	46
3.2.2 Database file format	50
3.3 Retrieval arithmetic	52
3.4 Summary	54
Chapter 4 System and Results	55
4.1 Function design	55
4.2 Human-computer interaction	58
4.3 Result and Analysis	62
Chapter 5 Conclusion and Future Work	67
5.1 Conclusion	67
5.2 Future Work	68
Appendix	69
Reference	71
Publications	75
Acknowledgements	77

第一章 绪论

1.1 哼唱识别与音乐检索系统概述

随着数字化技术的发展，日常生活与计算机的结合越发紧密，越来越多的事物需要在计算机中予以表示。从古至今，音乐都是人们最为喜爱的娱乐活动之一，随着 WAV 和 Mp3 等音乐格式的诞生，音乐已经能够储存于计算机之中，并实现了数字化。但是，只有这样的格式并不利于人们查找和检索音乐，特别是当用户以自己的哼唱来检索音乐时，往往难以实现高效而准确的检索。

所谓的音乐检索系统主要实现的是基于内容的音乐检索^[1, 2, 3]。基于内容的音乐检索技术（Content-Based Music Retrieval）是研究如何利用音乐信号的幅度、频率等物理特征，响度、音高、音色等听觉特征，节奏、旋律、调式、情绪等语义特征来实现音乐信息检索的一门新兴技术和学科。^[4]

音乐检索系统需要用户输入可用于检索的旋律。对于用户以哼唱来检索音乐时，还需要考虑到如何将哼唱转换为可以检索的关键字和查询信息，即将哼唱以检索为标准数字化。哼唱识别就是实现这种数字化的一种技术。哼唱识别是将录制好的哼唱文件转换为类似 Midi 的有序的音高和时值序列。这种音高和时值序列可以直接用于检索 Midi 音乐、数字化乐谱以及经过预处理的 Mp3 音乐数据库，以实现更为准确的检索。

1.2 研究的目的和意义

1.2.1 研究的目的

本研究旨在实现一个主要用于音乐检索系统的哼唱输入系统。该系统具备有效的音符划分和音高识别功能及友好的人机交互界面。系统的研究提出一个新的音乐检索预处理的研究方向。同时，在此基础上讨论与该哼唱输入系统相匹配的基于内容的音乐检索系统——哼唱检索系统（QBH）的实现。

1.2.2 研究的意义

在社会高速发展的今天，音乐作为一种大众娱乐，受到了广大民众的重视。每年，单从国内就有上千张音乐唱片诞生。可是，有这么多的音乐，却不是每个人都能知道甚至熟知的，这就需要有一个有效的音乐检索系统来帮助人们查找那些似曾相识却不明出处的音乐。目前，大多数音乐检索系统都是基于文本数据（歌名，歌词），而音乐的旋律才是用户最为敏感的。因此，基于内容的音乐检索系统才是音乐检索系统发展的主要方向。

基于内容的音乐检索系统以旋律为关键字检索。这种旋律关键字可以通过Midi 键盘输入，数字化乐谱和用户哼唱得到。但前两者方法对用户的乐理知识要求太高，只适用于专业从事音乐的用户，不适合推广；哼唱作为一种常见的娱乐方式，才是绝大多数人都能便捷使用的输入方法。由此，哼唱检索系统的研究又是基于内容的音乐检索系统的重中之重。

正如人们可以逐词或逐字通过语音将一段话输入计算机那样，那些可以哼唱某首歌曲的人也能通过本文研究的哼唱转换MIDI系统，将其心目中期望的歌曲旋律通过哼唱输入计算机，并以MIDI形式记录在计算机里。这就是哼唱识别技术所要实现的功能。成功的哼唱旋律辨识技术可直接提高哼唱检索系统的命中率。实际上，哼唱旋律辨识技术已经成为QBH系统研究中的一个重要问题。

1.3 相关研究概况和问题的提出

Ghias^[5]等，于1995年展示了第一个QBH系统，该系统采用时域上的自相关法提取音高，根据音高前后的变化，用S(same)、U(up)和D(down)三个字母来表示旋律的轮廓。但他们的系统并没有考虑到旋律的节奏。Lu^[6]等不仅考虑了旋律的轮廓，还考虑到音调的间隔及长度，提出了将这些信息组合，用于检索的新方法。他们同时采用了层次化搜索来提高检索的效率。Zhu^[7]等采用了动态时间规整（DTW）技术，将哼唱与音乐数据库直接比较。上海交通大学的李扬^[8]等，提出了一种近似旋律匹配的线性对齐匹配法，并且实现一个哼唱检索系统的原型。该算法并非基于近似符号串匹配、统计模型或者特征空间，而是根据相近旋律的音高轮廓在几何上的相似性，将音高和节奏特征一并考虑所设计而成的全新算法。浙江大学的冯雅中等，提出了一种哼唱检索音乐的启发式层次化方法^[9]，该方法可以减小检索的复杂度。中科院的吴晓等^[10]提出了一种基于三层旋律框架

的哼唱检索系统。台湾清华大学的张智星等^[11]为哼唱检索引入了渐进过滤器框架。Yang^[12]等使用一个 BP 神经网络将旋律转换为检索关键字用于检索。

在哼唱检索应用领域,当前效果很好的有“Mi Do Mi”网站提供的在线哼唱检索系统。该网站提供给用户上传自己哼唱的功能,上传后的哼唱提取出旋律特征,集合成为一个哼唱数据库。当用户使用自己的哼唱来检索音乐时,是用自己的哼唱去检索网站中已经存在的一个哼唱的旋律特征。经研究发现,该网站的检索成功率较高,但是成功的前提是哼唱者哼唱的节奏与目标歌曲是大致匹配的,如果节奏错误,即使哼唱的音高完全正确,也很难找到正确的匹配结果。也就是说该系统主要考虑的是以哼唱节奏来检索哼唱节奏。

国内外从事 QBH 系统研究的学者,也不同程度地对哼唱旋律辨识技术进行了各种有益的探索,并取得了很好的成绩。Lin^[13]等引入了 Score Predictive Model (SPM),该模型采用了预测分数函数来优化 HMM 和 DTW 的结果,从而获得更有效的音符切分。Krigel^[14]等在基于 HMM 的音乐翻译中,使用了过渡模型、无重复语法和约束状态模型,获得较好的成果。Ryynänen^[15]等在单声部音乐翻译中,采用了概率音符时间模型和音乐学模型;音乐学模型的引入为哼唱识别提供了一个新的研究思路。西北大学的王小凤^[16]提出了一种新的音乐旋律轮廓提取算法。

在旋律提取应用领域,比较知名的有 WIDI Recognition System (简称 WIDI) 系统。WIDI 系统的主要用途是音乐识别。它可以将波形文件识别为一个音符符号的序列,即 MIDI 文件。MIDI 文件与波形文件相比具有文件小,编辑器强等优点。WIDI 系统的缺点则是识别出的音乐效果与设备有关,并且不是所有声音都可以正常地转换为符号,比如人声。当用户使用 WIDI 系统处理人声哼唱时,可能可以大致听出音乐的主旋律走向,但是 WIDI 并没有将哼唱中各个基音对应的泛音过滤掉,造成获得的结果嘈杂混乱,完全无法用于哼唱检索。

虽然音频旋律到乐谱的自动转换(Automatic Melody Transcription)的研究已经有很长一段历史了,但是人声哼唱旋律的自动辨识技术目前还无法成熟到可进入实用阶段。这主要有二个方面的原因。其一是系统因技术原因而导致对哼唱旋律的误识。其二就是哼唱者本身没能正确地哼唱出他(或她)所期望的旋律。现有哼唱旋律辨识技术的研究成果多集中于前者。例如,哼唱基频音高的自动提取

及哼唱片段中的一个具有唯一确定的听觉音高且为单音节哼唱发音事件(对应哼唱旋律中的一个音符)的边界如何自动确定等等^[17, 18],而后者却较少有人关注。尽管不少人都能够记下并且通过自身的听觉来辨识心目中某首歌曲的旋律,可是人的哼唱能力并不像人的言语能力哪样普及。如果我们把十二平均律作为界定哼唱旋律中二个音之间音程的尺度,那么可以发现,并非每个人都能够把一段旋律准确无误地哼唱下来。有些哼唱者,在哼唱时会跑调,但自己却全然不知。有些哼唱者虽然能将整首哼唱旋律稳定在一个基调上,但仍会把某些音唱偏差半音音程或甚至更大。而这些偏差在旁人来说有些可能还听不出来。有些哼唱者还不能保持一个恒定的哼唱速度。如此种种,都说明一个问题,人的哼唱所展示的旋律常常和心目中期望的旋律有一定距离。

毫无疑问,任何一个全自动旋律辨识系统都难以从不准确的哼唱中直接获取一个准确表征哼唱者心目中期望旋律的符号乐谱表达或相应的类 MIDI 形式的旋律表达。欲使系统能从“并非完全正确”的哼唱中获取哼唱者心目中所期望的正确旋律表达,一个人机交互的旋律音高和音长校对过程是必需的。

1.4 本论文创新点

通过与当前相关领域研究现状对比,概括本论文研究之创新点如下:

(1) 文章提出了一种基于能量的多层次音符分割方法。当前的基于能量音符分割多为单层次的划分。单层次划分仅使用一种方法划分进行一轮划分,不能充分考虑到音符能量变换的各种情况。多层次划分可以对不同情况使用不同的划分方法,高层次的划分还可以在低层次的基础上处理,减少额外干扰因素。

(2) 目前研究领域内尚无基于乐理的音高识别技术。本文提出了一种新的基于乐理知识的音高识别方法,即基于倍音列的音高计算模型。通过倍音列理论的运用,不需要对哼唱信号预先进行滤波处理就可得到较高正确率的基音识别率,使得音高的识别和计算更加高效。

(3) 哼唱旋律的识别技术一般只实现到识别为止,本文在识别之后提出了通过调式规整音高的方法,作为识别的补充。该方法根据歌曲音乐调式的特性,自动试探歌曲的调式,通过得到的调式,规整已经识别出的音高,使得识别出的音高更佳和谐,也一定程度上提升了系统自动识别音高的准确性。

(4) 除上述音高规整方法外, 本文还讨论和实现了一种用户参与互动的哼唱输入方式。目前还尚未有一种用于哼唱旋律识别结果调整的人机互动系统。论文针对哼唱者的“哼唱误唱”, 专门设计一种基于人机交互的哼唱转换类似 MIDI 的音高和时值序列的系统。它可以帮助哼唱者通过哼唱及一个人机交互过程将其心目中旋律片段识别出来并用合成钢琴音色播放以再现旋律, 更容易表达出使用者心中期望的旋律。系统给出大致正确的音符划分和识别的音高, 用户可通过系统功能, 调整错误的音符划分和音高。通过互动, 系统得到一个更为准确的音高和时值序列, 为下一步的检索提供了一个很有效的检索关键字信息。

(5) 提出了一种专门用于哼唱检索系统的数据库构建方法和与之对应的基于哼唱旋律识别技术的哼唱检索算法。该数据库的构建是基于哼唱旋律识别系统的识别特点设计, 简便实用; 构建方法包括人工乐谱特征录入格式设计, 数据库存储格式设计和乐谱特征转换为数据库存储格式的方法。该检索算法有较好的健壮性, 比较适合已经识别出哼唱的较为准确的音高和时值序列的情况。

(6) 目前的哼唱旋律识别实验无统一评价标准, 本文提出了一种有效的哼唱旋律识别实验评判标准, 该标准能更加合理的表现出哼唱旋律识别技术的优劣。

1.5 论文篇章结构

本论文共分五章。第一章是绪论, 介绍了哼唱识别和基于内容的音乐检索、本课题的研究目的和意义、国内外的相关工作、和本论文的主要创新点。第二章按照哼唱识别处理的顺序分别介绍了哼唱识别技术两个主要步骤和一个附加步骤。第二章第一节研究了识别的第一步音块切分技术; 第二节介绍了第二步音高识别技术。第三节介绍了附加步骤音高规整技术。第三章探讨了基于哼唱旋律识别技术的哼唱检索系统的设计。第四章介绍了哼唱识别及检索系统的实现和实验结果分析和评估。第五章对全文做出总结, 并对未来的工作提出建议和展望。

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库